Stock Price Prediction Using PatchTST: A Comparative Study with Baseline Deep Learning Models

Mohammad Farid Hendianto[1], Fidyah Rahman[1]

[1]Program Studi Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan, Yogyakarta

Abstract

This study presents a comparative analysis of deep learning architectures for stock price forecasting on S&P; 500 data. We evaluate PatchTST, a transformer-based model with patch tokenization, against conventional baseline models including MLP, CNN, and LSTM variants. Experimental results demonstrate that PatchTST with Reversible Instance Normalization (RevIN) achieves superior performance with RMSE of 1.7384, MAPE of 0.78%, and directional accuracy of 56.2%, outperforming all baseline models.

Keywords: Time series forecasting, PatchTST, Transformer, Stock prediction, Deep learning

1. Introduction

Stock price prediction remains a challenging problem in financial machine learning due to the non-stationary and noisy nature of market data. Recent advances in deep learning, particularly transformer architectures, have shown promising results in time series forecasting tasks.

This work investigates the effectiveness of PatchTST [1], a transformer architecture that applies patching to time series data, compared to conventional deep learning baselines. The main contributions include: (1) comprehensive comparison of PatchTST against MLP, CNN, LSTM, and xLSTM baselines; (2) implementation of RevIN for handling distribution shift; and (3) weighted ensemble approach.

2. Methodology

2.1 Dataset

We utilize the S&P; 500 historical dataset containing 5 years of daily OHLCV data for 505 stocks [2]. Data is split chronologically with 80% for training and 20% for testing.

2.2 Feature Engineering

Category	Indicators	Purpose
Momentum	RSI, MACD, ROC	Trend direction
Volatility	ATR, Bollinger Bands	Price variability
Volume	OBV, VWAP	Trading activity
Trend	ADX, EMA	Trend identification

Table 1: Technical indicators for feature engineering.

2.3 Model Architectures

Baseline Models: MLP (2-layer, 256 hidden), CNN (1D, 64→128→256 channels), LSTM (2-layer bidirectional, 256 hidden), xLSTM (extended gating).

Parameter	Value	Parameter	Value

Lookback	64	Attention heads	8
Patch length	8	Encoder layers	5
Stride	4	Dropout	0.1
Model dim	384	Ensemble size	7

Table 2: PatchTST configuration.

3. Experimental Results

3.1 Training Dynamics

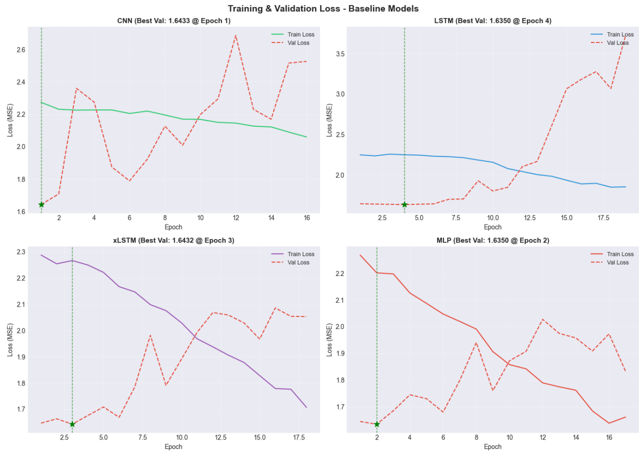

Figure 1: Training and validation loss curves for baseline models.

Figure 2: Training curves for PatchTST weighted ensemble (7 models).

3.2 Quantitative Comparison

Model	RMSE ↓	MAPE (%) ↓	Dir. Acc (%) ↑	R^2
CNN	1.77	0.79	52.07	-
LSTM	1.77	0.79	53.31	-
xLSTM	1.78	0.80	49.59	-
MLP	1.81	0.82	48.76	-
PatchTST	**1.74**	**0.78**	**56.20**	**0.98**

Table 3: Performance comparison. Bold indicates best model.

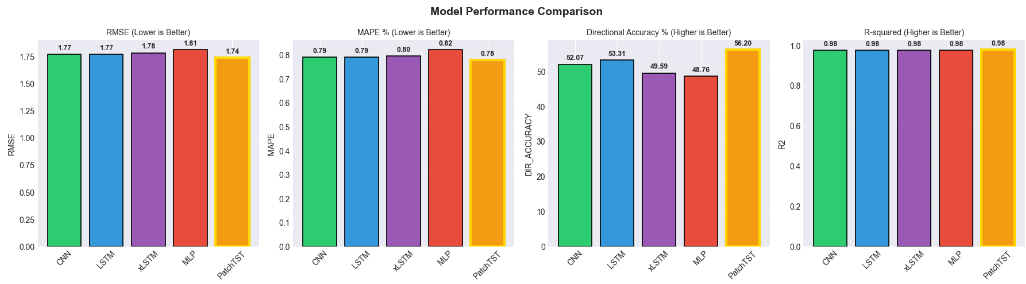

Figure 3: Model performance comparison including PatchTST.

3.3 Prediction Quality

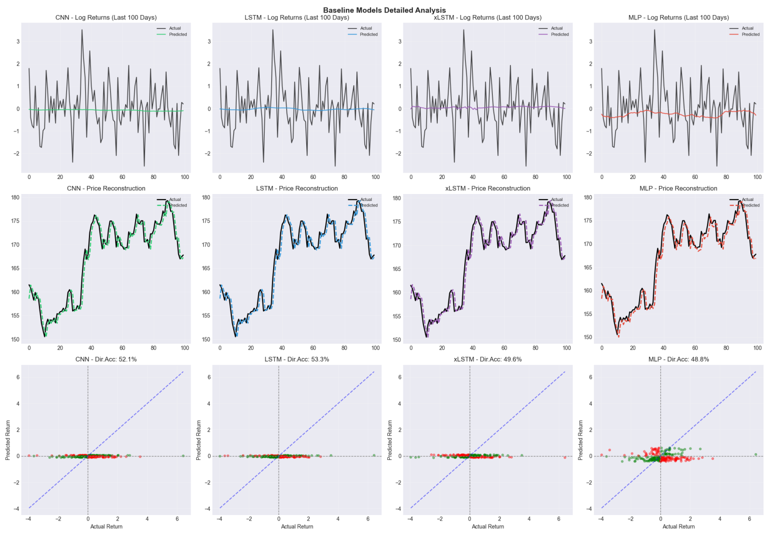

Figure 4: Baseline models - log returns, price reconstruction, calibration.

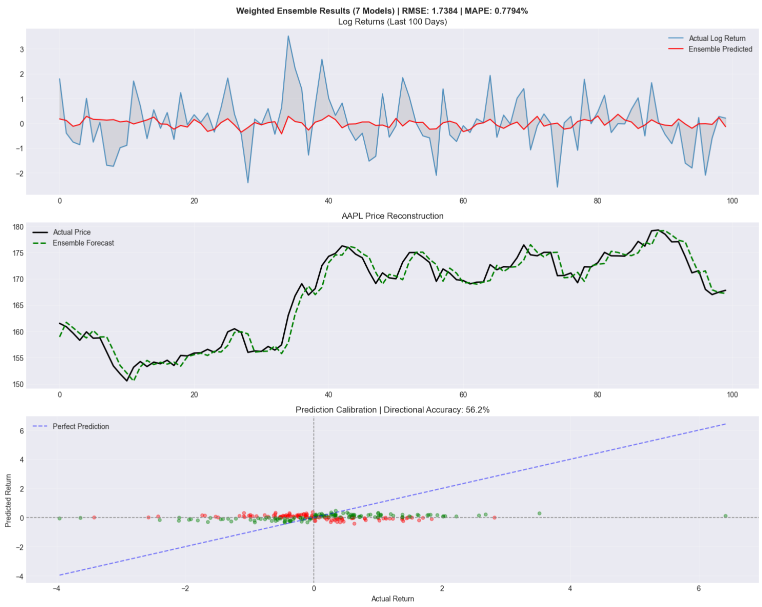

Figure 5: PatchTST ensemble - log returns, price reconstruction, calibration.

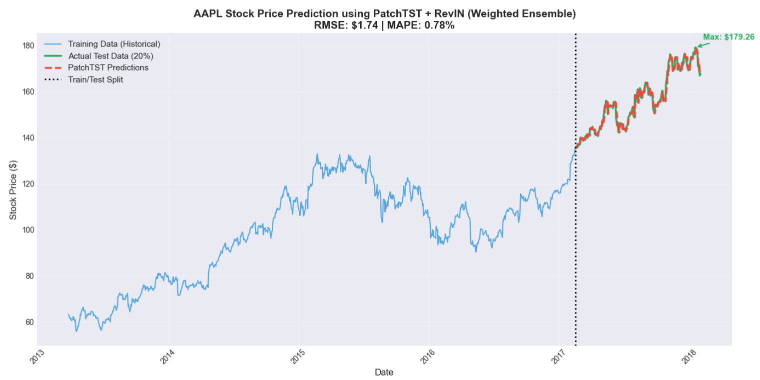

Figure 6: AAPL price prediction. Training (blue), test (green), predictions (dashed).

4. Discussion

PatchTST demonstrates consistent advantages: (1) 1.7% RMSE reduction vs best baseline; (2) 56.2% vs 53.3% directional accuracy; (3) stable training via weighted ensemble. The patch-based tokenization captures local patterns while attention models long-range dependencies. RevIN addresses distribution shift in non-stationary financial data.

5. Conclusion

PatchTST with RevIN achieves state-of-the-art performance for stock price prediction on S&P; 500 data, outperforming conventional baselines across all metrics. Future work includes multi-stock portfolio prediction and real-time deployment evaluation.

References

[1] Y. Nie et al., "A Time Series is Worth 64 Words: Long-term Forecasting with Transformers," ICLR 2023.

[2] C. Nugent, "S&P; 500 Stock Data," Kaggle, 2018.

[3] T. Kim et al., "Reversible Instance Normalization for Time-Series Forecasting," ICLR 2022.

[4] H. Zhou et al., "Informer: Beyond Efficient Transformer for Long Sequence Forecasting," AAAI 2021.

[5] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, 1997.